

**TOPIC IDENTIFICATION AND USE THEREOF IN INFORMATION
RETRIEVAL SYSTEMS**

Abstract

5 A technique to determine topics associated with, or classifications for, a data corpus uses an initial domain-specific word list to identify word combinations (one or more words) that appear in the data corpus significantly more often than expected. Word combinations so identified are selected as topics and associated with a user-specified level of granularity. For example,

10 topics may be associated with each table entry, each image, each sentence, each paragraph, or an entire file. Topics may be used to guide information retrieval and/or the display of topic classifications during user query operations.